JSLHR

Research Article

The Influence of Reading on Vocabulary Growth: A Case for a Matthew Effect

Dawna Duff,[a] J. Bruce Tomblin,[a] and Hugh Catts[b]

Purpose: Individual differences in vocabulary development may affect academic or social opportunities. It has been proposed that individual differences in word reading could affect the rate of vocabulary growth, mediated by the amount of reading experience, a process referred to as a *Matthew effect* (Stanovich, 1986).
Method: In the current study, assessments of written word–reading skills in the 4th grade and oral vocabulary knowledge collected in kindergarten and in the 4th, 8th, and 10th grades from a large epidemiologically based sample (n = 485) allowed a test of the relationship of early word-reading skills and the subsequent rate of vocabulary growth.
Results: Consistent with the hypothesis, multilevel modeling revealed the rate of vocabulary growth after the 4th grade to be significantly related to 4th-grade word reading after controlling for kindergarten vocabulary level, that is, above average readers experienced a higher rate of vocabulary growth than did average readers.
Conclusions: Vocabulary growth rate differences accumulated over time such that the effect on vocabulary size was large.

There are large differences between individual children in their vocabulary knowledge on school entry (e.g., Hart & Risley, 1995), and these differences in vocabulary extend into the school years. For example, Biemiller and Slomin (2001) reported that in the second grade, children at the lowest quartile for vocabulary had approximately half the number of known words compared to students in the top quartile. Furthermore, according to the Matthew effect model proposed by Stanovich (1986, 2000), those individual differences in vocabulary may even increase over time. The term *Matthew effect* refers to a biblical text and was originally proposed to describe the progress of scientific research careers (Merton, 1968) in which advantages and disadvantages accumulate, so that the rich get richer and the poor get poorer. In terms of reading, the general premise of the Matthew effect model is that individual differences in reading skill (broadly conceived) could accumulate over time (Stanovich, 1986, 2000) so that a child's initial reading level would be positively related to his or her rate of growth in a reading skill. This pattern, in which growth rates differ across skill levels even while absolute skill levels increase for all, is considered a relative Matthew effect (Rigney, 2010). Accumulating advantages and disadvantages, of course, are only one possible developmental pattern. A compensatory model would predict that initial reading level would be negatively related to rate of growth in reading skill so that differences in reading skill would decrease over time, effectively the opposite of a Matthew effect (Pfost, Hattie, Dorfler, & Artelt, 2014). A third possibility would be a stable achievement pattern, with high and low skill readers having the same rates of growth over development (Pfost et al., 2014).

This study concerns one specific prediction of the Matthew effect model, namely, that reading skill in general, and word reading skill in particular, could be related to the rate of vocabulary growth. Vocabulary skill is strongly related to a variety of academic, vocational, and social outcomes (Dollinger, Matyja, & Huber, 2008; Gertner, Rice, & Hadley, 1994; Rohde & Thompson, 2007). The veracity of this prediction of the Matthew effect model is significant because it could help guide interventions for children at risk of poor vocabulary development. The current study includes children sampled from a large epidemiologic study, which includes children with language impairments and cognitive impairments.

The Role of Reading in Vocabulary Development

The prediction that reading skill could be associated with rate of vocabulary growth is based on the premise that reading development could potentially have a significant

[a]University of Iowa, Iowa City
[b]Florida State University, Tallahassee

Correspondence to Dawna Margaret Duff: dawnaduff@gmail.com

Editor: Rhea Paul
Associate Editor: Marleen Westerveld

Received November 13, 2013
Revision received May 17, 2014
Accepted February 15, 2015
DOI: 10.1044/2015_JSLHR-L-13-0310

Disclosure: The authors have declared that no competing interests existed at the time of publication.

effect on a child's exposure to novel words. In fact, there is empirical evidence that, for older children and adults, much learning of new words occurs through exposure to written texts (Nagy, Herman, & Anderson, 1985; Sternberg, 1987). Because print material generally contains many more low frequency words than does spoken language (Cunningham, 2005), reading text can provide key opportunities for advancement in vocabulary development. We predict that word learning through reading will affect vocabulary as measured on both oral and written tasks because words learned through reading text will be at least partially available to the individual for both written and oral language use (Nelson, Michal, & Perfetti, 2005).

However, exposure to novel words in text does not occur uniformly throughout reading development. Prior to formal literacy instruction, children are clearly acquiring novel vocabulary through exposure to oral language. During early reading development, children rarely confront words in print that are not already present in their vocabulary, so much of the lexical knowledge of words, especially phonological and semantic representations, will be derived from oral language experience. As children become more proficient readers and advance to more complex print material, they are more likely to confront words during reading that they have not been exposed to via listening. This transition likely occurs around the third or fourth grade for many students (Chall, 1987). Biemiller (2005), for example, reported that, from the third grade onward, but not in earlier grades, 95% of children could read more words than they could explain.

The Existence of a Matthew Effect for Vocabulary

According to one of the predictions of the Matthew effect model, vocabulary development after the third or fourth grade would be affected by reading ability and the associated reading experiences enabled by these reading skills. This study investigates a rather straightforward prediction with respect to vocabulary development and reading during the middle grades and high school. We predict that better readers during this time will have a greater likelihood of confronting novel, low-frequency words than will weak readers and that this will affect the rate of vocabulary growth. This prediction is predicated on the notion that strong readers will engage in more reading activities than will weak readers. This assumption is consistent with Stanovich's (1986) proposal that the volume of reading experience is the key mediating variable between reading skill (broadly conceived) and vocabulary, with cumulative advantages occurring due to "the effect of reading volume on vocabulary growth, combined with large skill differences in reading volume" (p. 381). There is empirical evidence to support the assumption that reading skill and the amount of reading experience are strongly associated. For example, Allington (1983) reported that strong 1st-grade readers read three times as many words during reading instruction as do weak readers. Nagy and Anderson (1984) suggested that a motivated middle-school student might read 100 times more words a year in the classroom than a less skilled or motivated student. With respect to reading for pleasure, Juel (1988) reported that average and strong readers in the third and fourth grades read at home more times per week than did weak readers, and Martin-Chang and Gould (2008) reported correlations between reading speed (words per minute) and personal reading experience in undergraduate students.

A small number of studies have previously investigated a Matthew effect with vocabulary as an outcome variable. Aarnoutse and van Leeuwe (2000) reported that weak readers showed larger effect sizes in vocabulary growth than did strong readers in early elementary grades, thus leading the authors to question a Matthew effect of reading on vocabulary. Vocabulary was measured in a written format; thus, reading ability could have confounded the measure of vocabulary. In contrast, Cain and Oakhill (2011) reported that readers who had weak reading comprehension skills showed lower rates of vocabulary growth between the ages of 8 and 16, compared with good comprehenders, and concluded that there was a Matthew effect for reading skill on vocabulary. In this case, vocabulary was measured via both word reading and listening vocabulary. In a similar manner, Kempe, Eriksson-Gustavsson, and Samuelsson (2011) reported evidence of a Matthew effect on the growth of vocabulary in the 1st to third grades, as measured orally using the Wechsler Intelligence Scale for Children–Third Edition (Wechsler, 1991). In addition, Stothard, Snowling, Bishop, Chipchase, and Kaplan (1998) reported a decrease in scores on the British Vocabulary Scale (Dunn, Dunn, Whetton, & Pintilie, 1982) between ages 8 and 15 for children who had been classified as having persistent specific language impairment and general delay, but not for children whose language was within the expected range or for children whose early language concerns had resolved by age 5 years. None of these studies used developmental scaling to equate item difficulty across different age groups. Further, none of the afore-mentioned studies controlled for rate of vocabulary learning prior to literacy instruction. It is reasonable to expect that the various factors that contribute to these individual differences in word learning in early life might continue to exert effects on word learning when reading. There are, in fact, substantial differences in word-learning achievement in prereaders (e.g., Hart & Risley, 1995), which would affect the level of vocabulary knowledge when children start to learn new words through written language. Furthermore, these individual differences in word-learning skills would be expected to covary with reading skill, given the substantial overlap between disorders of word reading and of language skills (Catts, Adlof, Hogan, & Ellis Weismer, 2005). In order to examine the specific effect of reading experience on vocabulary, it would seem wise to control for the child's general word-learning achievement. Thus, the evidence for a Matthew effect on vocabulary is mixed and is possibly confounded by word-learning abilities in general.

The above discussion concerns the effect of reading skill on vocabulary growth, which is only one prediction of the Matthew effect model. Other predictions of the model have also been tested, with similarly equivocal results (Pfost et al., 2014). Some studies report data that support a Matthew effect for reading ability (Juel, 1988), but others report a stable achievement pattern (Aarnoutse & van Leeuwe, 2000; Catts, Adlof, & Fey, 2003; Scarborough & Parker, 2003; Shaywitz et al., 1995) or a compensatory effect (Parrila, Auonola, Leskinen, Nurmi, & Kirby, 2005; Shaywitz et al., 1995). The diversity of findings in these studies is undoubtedly related to the wide variety of outcome variables and ages of readers as well as to the characteristics of the sample group and study methodologies. Indeed, some studies do report different conclusions on the basis of the outcome variable studied (Bast & Reitsma, 1998; Shaywitz et al., 1995), the subgroup of children looked at (Jacobson, 1999; Morgan, Farkas, & Hibel, 2008; Phillips, Norris, Osmond, & Maynard, 2002; Stothard et al., 1998), and even the language in which children were learning to read (Parrila et al., 2005). In addition, a recent meta-analysis (Pfost et al., 2014) concluded that the psychometric properties of the measures were also important: studies using measures without floor or ceiling effects and with good reliability were more likely to report the presence of a Matthew effect. As a final consideration, the populations studied may have differed in amount or kind of intervention received. Hence, although the Matthew effect model has been a very helpful framework for researchers, educators, and clinicians alike, evidence for it has remained elusive (Pfost et al., 2014; Scarborough & Parker, 2003).

Where a Matthew effect is reported, there is more than one possible pattern because the effect of initial reading skill on subsequent growth rates may not necessarily be the same across the continuum of reading skill (Protopapas, Sideridis, Mouzaki, & Simos, 2011; Rigney, 2010). On the one hand, strong readers might show increasing gains relative to average readers at the same time as weak readers show decreasing gains relative to average readers. We refer to this as a *two-sided Matthew effect*. On the other hand, strong readers could show increasing gains relative to average readers, whereas weak readers have gains similar in size to those of average readers. The reverse of this pattern is also possible in which weak readers show slower growth rates than average readers without strong readers showing faster growth rates (e.g., Morgan et al., 2008). These last two possibilities have been termed *one-sided Matthew effects* (Morgan et al., 2008), and we describe them as such.

It is clear that the selection of outcome and predictor variables is of critical importance in tests of a Matthew effect. Stanovich's (1986) proposal about reading and vocabulary considered reading in a broad sense. In this study, word reading (of nonwords and single words) is used to operationalize reading skill. The rationale for using word-reading skill as a predictor variable is simply that is expected to be less confounded with vocabulary than reading comprehension scores would be because reading comprehension and vocabulary scores are highly correlated (e.g., Pearson, Hiebert, & Kamil, 2007). The use of word-reading scores therefore allows for a clearer interpretation of the data. Likewise, vocabulary can be defined in different ways, including across receptive and expressive dimensions. The data set used in this study has been previously analyzed for receptive/expressive dimensionality using revised modified parallel analysis and confirmatory factor analysis (Tomblin & Zhang, 2006). This analysis concluded that the measures used in the study "are not likely to be able to reflect reliable differences within individuals with respect to receptive and expressive modalities" (p. 1206). Hence, despite the use of different tasks in receptive and expressive vocabulary measures in this study, the latent trait measured does not seem to be different. Therefore, in this study, vocabulary skill is operationalized as a composite score, including both receptive and expressive measures.

Study Questions

This study will test the specific prediction that rate of vocabulary growth is related to reading skill by examining the growth in oral vocabulary in an epidemiologically based sample between the fourth and 10th grades among children with a wide range of reading abilities, established at the fourth grade. The first specific question of this study is, is there evidence that fourth-grade word-reading skill is related to the rate of change of vocabulary growth between the fourth and 10th grades after accounting for individual differences in the level of vocabulary acquisition prior to reading instruction? In the current study, vocabulary skill in kindergarten is used as a measure of these individual differences in word learning prior to formal reading instruction. The hypothesis, based on Stanovich's (1986) model, is that that there will be a relationship between fourth-grade reading skill and the rate of vocabulary growth in the years between the fourth and 10th grades.

The second specific question of this study is, if there is a relationship between reading skill and vocabulary growth, is this relationship the same for both strong and weak readers? In other words, if a Matthew effect exists, is it a one-sided or a two-sided Matthew effect? There was no hypothesis for the second question in this study because no previous research has addressed this specific question and there might be some reason to expect either a two-sided or a one-sided Mathew effect. As Shefelbine (1990) pointed out, readers with lower initial vocabulary knowledge will necessarily have an impoverished semantic context for inferring new word meaning, which might lead to lower rates of vocabulary growth. On the other hand, those same readers are less likely to encounter ceiling effects because any given text is more likely to include words that are novel to them. This argument made by Shefelbine (1990), however, concerns the effect of initial vocabulary skill on vocabulary growth. This is in contrast to the current study, which addresses the relationship of reading skill and vocabulary growth.

Method

Sample

The data analyzed in the current study were drawn from a sample of 604 participants who originally took part in an epidemiologic study of language impairment (Tomblin et al., 1997). The original epidemiologic sample participated in the 1993–1994 school year and consisted of 7,218 kindergarten students, representing all available kindergarten students who were monolingual English speakers in selected schools in rural, urban, and suburban areas in Iowa and Illinois. In this initial sample, a stratified cluster sample was used, with stratification by residential setting and cluster sampling according to school (Tomblin, 2014). All students who failed the initial screening were given a diagnostic battery of language and cognitive measures, as were a representative sample of students who passed the screening, such that the group who passed the screening battery and the group who did not were of equal size. Each of these participants was recruited to be part of the longitudinal study, and all who consented became participants in the longitudinal study. All children who completed the longitudinal study and for whom vocabulary scores were available (n = 485) were included in this sample. The children in fourth grade averaged 10.0 years (SD = 0.40), and in the 10th grade, they averaged 15.8 years (SD = 0.37).

The original sample of 485 children contained an oversample of children with poor language abilities. Because this oversampling was applied systematically to a population sample, it was possible to derive a weighting system that adjusted for this; that is, scores were weighted by multiplying each child's score by a constant that was equal to the expected prevalence of that diagnostic category and gender divided by the actual prevalence of those children in the sample. In this manner, children with poor language received proportionally less weight in the analyses than did children who showed typical language, a weighting procedure has been described in other published work involving the sample (Catts et al., 1999, 2005). The resulting sample of 485 children contained an equal proportion of boys and girls (50% of each). The distribution of the mothers' educational level was as follows: 4% had less than a 4-year high school education, 28% had a high school diploma, 41% had postsecondary education, 15% were college graduates, and 12% had postgraduate education.

As shown in Table 1, standardized language measures in the fourth grade and performance IQ measured in the second grade before weighting were below the expected population means. However, after weighting the samples, they are very representative of a normal population. The current study uses data from all participants, with weighted measures for all analyses.

Tasks

All tasks were administered as part of a larger longitudinal study (for a complete description, see Tomblin & Nippold, 2014). Administration of tasks was standardized, and each examiner was given detailed training and monitoring by a data-collection manager, with a minimum of 5% of examination sessions scored blindly by both the examiner and the data collection manager to ensure consistency in scoring, as well as in administration (Tomblin, 2014). Scoring of all tasks was done relative to the child's age at the time of testing.

Vocabulary

For the present study, the following vocabulary measures were analyzed: in kindergarten, the Picture Identification and Oral Vocabulary subtests of the Test of Language Development–Primary: Second Edition (Newcomer & Hammill, 1988), and in older grades, the Peabody Picture Vocabulary Test–Revised (PPVT-R; Dunn & Dunn, 1981) as well as the Expressive subtest of the Comprehensive Receptive and Expressive Vocabulary Test (CREVT; Wallace & Hammill, 1994). Receptive vocabulary measures were picture identification tasks, and expressive vocabulary measures were definition-generation tasks. Each of these is a well-established standardized measure, and information about their validity, specific to this data set, has been published (Tomblin, Nippold, Fey, & Zhang, 2014).

Reading

For the present study, the following reading measures were analyzed: the Word Attack (WA) and Word Identification (WI) subtests of the Woodcock Reading Mastery Test–Revised (Woodcock, 1987), which involve reading nonwords (WA) and sight words (WI). These measures are considered to be reliable assessments of word-reading skill (Cooter, 1989).

Analysis

Composite Developmental Ability Scores

Analysis of growth in a cognitive ability such as vocabulary requires that the children's performance be scaled on a continuum across the developmental period of interest. The principal challenge for the creation of a developmental scale is that the ability of the children must be measured

Table 1. Descriptive statistics of sample with regard to oral language measures in the fourth grade and performance IQ obtained in the second grade.

Test (source; expected population mean)	Weighted mean (SD)	Unweighted mean (SD)
Wechsler Intelligence Scale for Children–Revised: Performance Scale (Wechsler, 1991; 100)	100 (15)	95.39 (14.80)
Peabody Picture Vocabulary Test–Revised (Dunn & Dunn, 1981; 100)	102.83 (15.52)	95.40 (16.07)
Clinical Evaluation of Language Fundamentals–Third Edition: Recalling Sentences (Semel, Wiig, & Secord, 1996; 10)	10.34 (3.1)	8.86 (3.14)
Clinical Evaluation of Language Fundamentals–Third Edition: Concepts and Oral Directions (Semel, Wiig, & Secord, 1996; 10)	10.34 (3.2)	8.50 (3.08)

by different items at different developmental time points; thus, the items need to be equated with each other in some meaningful way across development. In this study, developmental ability scores were computed using a Rasch model of item response theory (IRT). The resulting scores are often viewed as being well suited for growth curve modeling (O'Malley, Francis, Foorman, Fletcher, & Swank, 2002). Within IRT, the probability of an item being passed in a test is a function of the participant's ability level, the item's difficulty, as well as its discrimination and the probability of guessing. When calibrating (Mislevy & Bock, 1998), guessing can be set as a constant, and the probability for given items can be calculated from the administration of test items to participants. Thus, item difficulty could be calculated by holding examinee's ability constant. This is termed *item calibration* (Mislevy & Bock, 1998). Items that were administered across more than one grade level, and which had overall pass rates of between 10% and 90%, were used as anchors (Vale, 1986) for this item calibration. These anchor items were then used to calibrate item difficulty across age levels. For example, if Items 8 and 9 were given to fourth graders, and Items 9 and 10 were given to eighth graders, Items 8 and 10 can be calibrated via their overlap with Item 9 across grades. Table 2 provides a list of the specific items from the PPVT-R and the CREVT at each grade level used in the item calibration, resulting in Rasch-scaled vocabulary ability scores across the fourth to the 10th grades. The difficulty and the discriminating estimates for these items, along with estimates of expressive and receptive vocabulary ability for each examinee at each grade level, were computed using the computer program Bilog (Mislevy & Bock, 1998). Item parameters were determined using marginal maximum likelihood estimation. The 0 value on the scale was set for the average 6-year-old. Resulting ability scores provided a means of measuring the examinees' ability across time.

Weighted Scores

As described above, weighted scores were used in the analyses of this study to correct for the high rate of language and/or cognitive impairment. This weighted scoring procedure is possible because of the availability of data from the carefully sampled pool of participants in the epidemiological sample. This ensures that the data analyzed in this study are representative of the epidemiological sample, including children with and without a history of language impairment.

Table 2. Items from the Peabody Picture Vocabulary Test–Revised (PPVT-R) and the Comprehensive Receptive Expressive Vocabulary Test (CREVT) used in the item calibration.

Test items	Grade 4th	Grade 8th	Grade 10th	Number of items
PPVT-R				
80–83	X			4
85	X			1
86–94	X			9
138–151		X	X	14
CREVT				
2 and 3				2
4–10	X			7
11–15	X	X	X	5
16 and 17	X	X	X	2
18–24		X	X	7

Composite Scores

A composite score was derived for vocabulary for each participant, as discussed earlier. The composite was the mean of the developmental ability scores for receptive and expressive vocabulary. These composite scores were used to plot vocabulary growth curves.

In a similar vein, a composite score for word reading was calculated from the WA and WI scores at the fourth grade. A composite of these scores was used to incorporate the earlier developing skill of reading nonwords with the later developing skill of context-free word recognition (Tunmer & Chapman, 2012). Within the context of this study, these skills were used to index basic reading skills in the fourth grade. We expect that these skills are also indirectly indicative of the volume and variety of reading experience that these children will obtain after the fourth grade. This assumption is supported by a meta-analysis by Mol and Bus (2011) that indicated moderate correlations between print exposure and measures of WI and WA during elementary school years. Because the word-reading scores were part of the analysis at a single time point only, developmental scores were not required.

Multilevel modeling. Multilevel modeling was used to test the questions in this study, a method that is expected to yield comparable results to latent growth curve analysis (Chou, Bentler, & Pentz, 1998). Multilevel modeling of the weighted data in this study consisted of fitting each participant's vocabulary ability across the fourth, eighth, and 10th grades with parameters of intercept and linear slope. These parameters served as random effects in combination with a fixed effect of fourth-grade word reading as well as with the covariate of kindergarten vocabulary and their interactions with time (age) in a mixed model analysis using Proc Mixed software (SAS Institute, 2011).

The particular question of interest was whether the slope in vocabulary differed in accord with variation in fourth grade word reading. However, it could be argued that any association between word reading and vocabulary growth in later school years was merely because strong word learners become strong readers. To the extent that this is the case, the basis for the relationship would not be attributable to a special influence of reading on vocabulary. To address this, we also included the kindergarten vocabulary abilities of these children in this analysis as a covariate in this model on both the slope and the intercept. This provides a test of whether word reading is related to vocabulary after controlling for the children's word-learning achievement during the years prior to formal reading instruction. This was considered to be a direct test of the long-term relationship between word-reading skill and vocabulary development.

Because word reading was related to the rate of vocabulary growth, we computed the effect size in the form of f^2, which reflects the amount of variance in individual differences in vocabulary explained by fourth-grade reading after controlling for kindergarten vocabulary. This measure of the effect of reading on vocabulary growth concerns differences in slopes. A key feature of differential growth rates is that the individual differences accumulate over time; thus, the effect of the predictor variable—in this case, fourth-grade reading—on the outcome variable is likely to increase. Therefore, we measured the degree of association between fourth-grade word-reading ability and 10th-grade vocabulary after controlling for fourth-grade vocabulary ability.

Individual differences in vocabulary growth. Question 2 asked whether the effects of fourth-grade word reading on vocabulary growth were equally distributed across the range of word-reading ability. To do this, growth rates were contrasted between three groups of children categorized according to whether they had high, medium, or low fourth-grade word-reading ability. Vocabulary growth curves were plotted for participants with high word-reading skill (those who scored in the 80th percentile and above), middle word-reading skill (those who scored in the 40th–60th percentile range), and low word-reading skill (those who scored in the 20th percentile and below). We then used mixed modeling to contrast the middle group with the high and low groups with regard to growth rates.

Results

Vocabulary Growth Curves

As expected, the mean of the composite developmental ability scores for vocabulary showed an increase in vocabulary knowledge at each time interval, beginning in kindergarten (Figure 1). The average vocabulary score for the fourth-grade children was 2.26 (SD = 0.56) and was 3.59 (SD = 0.61) by the 10th grade. Figure 1 shows that the shape of the growth function was clearly nonlinear, with higher rates of vocabulary growth in early grades. However, between the fourth and 10th grades, the change was much more linear; thus, a linear model of vocabulary growth during this period of development was suitable. Figure 2 shows the mean vocabulary growth curves for readers with low, medium, and high reading skill in grade 4.

As anticipated, vocabulary ability upon school entry at kindergarten was correlated with vocabulary ability at fourth (r = .39, n = 485, p < .0001), eighth (r = .52 n = 485, p < .0001), and 10th grades (r = .55, n = 485, p < .0001). These correlations show that vocabulary ability at the onset of reading is associated with subsequent vocabulary ability, and therefore it is likely that there are factors influencing vocabulary growth in children that are not an outgrowth of their reading. In the subsequent analyses, the child's vocabulary ability in kindergarten will be used to represent these non–reading-related vocabulary learning skills and will be used as a covariate in order to better isolate later growth in vocabulary that is associated with reading ability.

Multilevel modeling using Proc Mixed software was used to test for differences in vocabulary growth across time where the child's age at testing was used to reference time. The results of this modeling are shown in Table 3. These results show that the mean vocabulary intercept (average vocabulary at age 9) before entering covariates (unconditional model) was 2.16 units of developmental ability score, and the mean growth rate was 0.24 points per year. Plots of the modeled values from a random sample of children at different levels of fourth-grade reading levels are shown (Figure 3). Because a linear model was used, these growth functions do not have the nonlinear quality of the data in Figures 1 and 2 but, otherwise, these modeled data are similar to the obtained data.

Figure 1. Distribution of developmental ability scores for vocabulary at each observational interval from kindergarten through 10th grade for all children in the longitudinal study

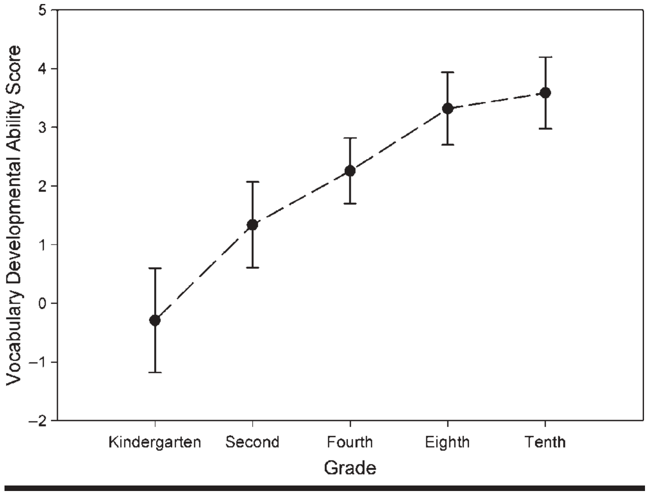


Figure 2. Vocabulary scores of participants grouped by word-reading skill (high-level readers, readers in the 80th percentile and above; midlevel readers, readers in the 40th–60th percentile range; low-level readers, readers in the 20th percentile or below).

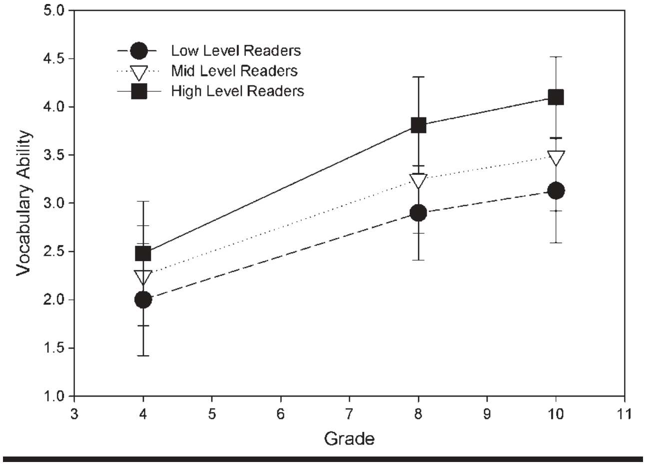


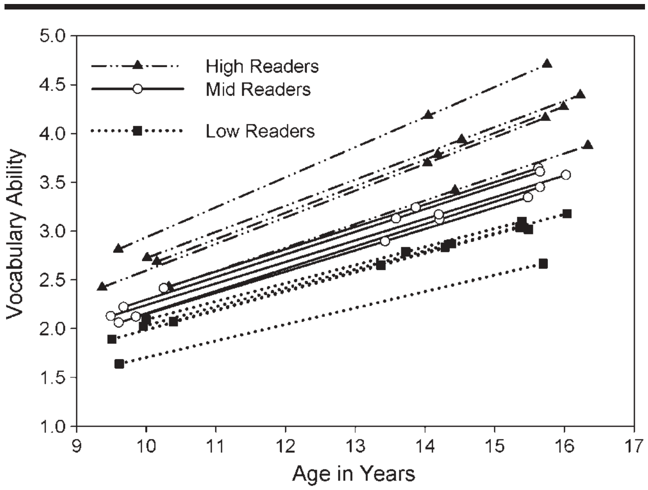
Table 3. Tests of random (Level 1) and fixed effects for vocabulary growth using kindergarten vocabulary as a covariate.

Model	Effects	Parameter	F value	p level
Unconditional				
	Intercept	2.162		
	Age	0.240	2446.87	<.0001
Conditional				
	Intercept	1.450		
	KV	0.146	19.60	<.0001
	4GR	0.008	17.97	<.0001
	Age	0.114	12.56	.0004
	Age × KV	0.016	6.64	.01
	Age × 4GR	0.001	15.88	<.0001

Note. KV = kindergarten vocabulary; 4GR = fourth grade reading.

Our first question concerned the degree to which fourth-grade reading is associated with vocabulary development (rate of change of vocabulary scores). This was tested within a conditional mixed model as also shown in Table 3. Table 3 shows that after controlling for their kindergarten vocabulary level, the rate of growth in vocabulary between the fourth and 10th grades was significantly associated with the children's fourth-grade reading level, $F(1, 485) = 15.88$, $p < .0001$. The parameter value for this effect was .001, which indicates that higher word reading ability in the fourth grade is associated with greater rates of vocabulary growth, which can be seen in Figure 2. It should be emphasized that this effect was estimated after the sources of variance concerned with kindergarten reading on overall vocabulary and change in vocabulary between the fourth and 10th grades had been entered into the model. Individual differences in kindergarten vocabulary also had a significant effect on the intercept of vocabulary score at the fourth grade, $F(1, 485) = 19.60$, $p < .0001$, and on the growth of vocabulary, $F(1, 485) = 6.64$, $p = .01$. Thus, children with higher vocabulary in kindergarten were likely to have higher vocabulary in the fourth grade and show greater rates of vocabulary growth after controlling for the effects of fourth-grade reading ability.

Figure 3. Linear growth functions for random samples of readers for readers in high (80th percentile and above), medium (40th–60th percentile range), and low (20th percentile or below) skill groups for word reading in fourth grade.



The slope parameter was found to be 0.001. This value represents the rate of change in vocabulary ability scores per year from the fourth through 10th grades that can be attributed to a 1-point change in fourth-grade word-reading skill. The average change in vocabulary ability across the 6 years was 1.38; however, some children changed by as much as 3.62 points and some actually declined by as much as −0.018. The relative size of the effect of fourth-grade reading ability on vocabulary growth compared with the overall growth in vocabulary can be represented using Cohen's f^2, which reflects the proportion of variance of a single variable within the context of a multivariate regression model. Using a method developed by Selya, Rose, Dierker, Hedeker, and Mermelstein (2012), we estimated that the individual differences in fourth-grade reading ability accounted for 8% ($f^2 = .08$) of the variance in vocabulary growth rates from the fourth to 10th grades. Cohen's f^2 of this size are generally regarded as small to medium in magnitude.

The effect size above reflects the degree to which fourth-grade word reading accounts for variation in the slopes of vocabulary growth across children. Although this effect size is somewhat small, we need to consider that differences in growth are likely to accrue over time. Thus, even small differences in growth rates can lead to substantial long-term effects in absolute vocabulary skill. In order to examine this cumulative effect of differential growth, we computed the magnitude of the effect of fourth-grade word reading at 10th-grade vocabulary after controlling for fourth-grade vocabulary. Thus, this reflects the gain in the effect of reading on vocabulary between the fourth and 10th grades. This resulted in an $\eta^2_{\text{partial}} = .26$. Eta squared values of this magnitude are viewed as large, and thus we can see that small to moderate effect size of differential vocabulary growth can result in a large effect given sufficient time.

We can also interpret the magnitude of this effect by comparing this effect size to something that is more familiar. In this case, we can compare the effect of mother's education level on vocabulary development during this same period. It is known that socioeconomic status is related to early vocabulary levels (e.g., Hart & Risley, 1995), and we would expect that the effect of maternal education would also extend into the school years. Using the same approach to compute f^2, we estimated the effect size for maternal education on vocabulary growth between the fourth and 10th grades and found that it was $f^2 = .08$. Thus, the effect obtained for fourth-grade reading ability on vocabulary is the same as that for maternal education.

The analysis above introduces the question of whether maternal education could be confounded with fourth-grade reading ability and whether this is the reason the effect sizes are similar. In this case, it could be argued that it is the child's home environment that explains the differential growth in vocabulary. However, in our test of reading effects on vocabulary between the fourth and 10th grades, we had controlled for kindergarten vocabulary, and one of the reasons for this was to control for socioeconomic factors that influence vocabulary growth. We tested this assumption by introducing both mother's education and kindergarten vocabulary in the same model and found that mother's education was not a significant predictor, $F(4, 474) = 0.24$, $p = .91$, of vocabulary growth after including kindergarten vocabulary. Thus, our inclusion of kindergarten vocabulary did effectively serve as a proxy variable for mothers' education.

The second question asked whether growth rates in vocabulary differed for the three groups of readers. This was addressed by performing a multilevel modeling analysis where the three groups of readers (high, medium, and low) were identified according to their fourth-grade word reading. Figure 2 is a plot of vocabulary growth functions for high-, medium-, and low-skill readers (as defined earlier). A pattern of divergence was shown. The significant effect of fourth-grade reading on growth confirms that differential growth in vocabulary exists in accord with fourth-grade reading; however, this differential could be concentrated in one region of reading ability. Contrasts in the group vocabulary growth between the low-level readers and the midlevel readers showed that the growth slope of the low readers was −0.02 ($SE = 0.0158$) lower than that for the midlevel group, which was not significantly different, $t(289) = 1.25$, $p = .21$. In contrast, the growth slope of the high-level readers compared with the midlevel readers was 0.04 ($SE = 0.0160$) higher, which was significantly different, $t(289) = 2.18$, $p = .03$. These results suggest that the association of fourth-grade reading ability and subsequent vocabulary growth varied somewhat depending on the reading level; that is, the effect was a one-sided, not a two-sided, Matthew effect. In this case, the readers in the upper 20 percentile showed divergence in vocabulary growth relative to those in the middle or low levels of the reading ability distribution.

Discussion

Relationship of Reading to Vocabulary Growth

The first specific question of this study was whether there was evidence that fourth-grade word-reading skill was related to the rate of change of vocabulary growth between the fourth and 10th grades after accounting for individual differences in vocabulary acquisition prior to reading instruction. Our results strongly support an association between word-reading ability and the rate of subsequent vocabulary growth as measured via an oral language task. It is quite unlikely, however, that word-reading ability in the fourth grade alone is sufficient to explain these results. Instead, we view our measure of fourth-grade reading ability as an indicator variable that is associated with reading-related activities of the children that unfolded between the fourth and 10th grades. These reading-related activities serve as the primary causes of vocabulary growth found in this study. We might add that the type of text the child is reading is also expected to be a variable because reading material that exposes the child to a wider range of vocabulary should also benefit vocabulary growth. Related to this point, Pfost, Dorfler, and Artelt (2013) reported that time reading narratives was much more predictive of vocabulary than was time reading newspapers, magazines, comics, or nonfiction. Thus, the results of this analysis are consistent with Stanovich's (1986) proposal as well as with that of Nagy et al.'s (1985) view that vocabulary growth during school years is largely due to incidental learning from written contexts. Given the importance of reading activity, we would ideally have measured these variables. Within this project, several measures of engagement in reading, such as author recognition, were collected but were found to be of questionable validity. However, other studies have shown an association between reading skill and the volume of reading experience (Allington, 1983; Martin-Chang & Gould, 2008; Nagy & Anderson, 1984). Nonetheless, although we assume that reading experience is a mediator of the relationship between word reading and the outcome of vocabulary growth, this mediation was not tested as part of this study. Therefore, the results of the current study do not allow us to draw conclusions about whether reading experience is, indeed, the mediator of the effect we found.

Once kindergarten vocabulary levels were accounted for, word reading in the fourth grade accounted for 8% of total variance in rates of vocabulary growth between the fourth and 10th grades. This means that the effect of word reading on vocabulary growth is not trivial. In fact, the size of the effect of word reading on vocabulary growth rates is comparable to the effect of maternal education on vocabulary growth rates during the same developmental period. When the impact of that rate difference is considered in terms of absolute vocabulary levels in the 10th grade, the effect is large.

As Stanovich (2000) stated, his 1986 article contains "many micropredictions and microtheories" (p. 150). Previous studies of other Matthew effects have reported variable results. There might be several reasons for these equivocal results for Matthew effects in previous studies. First, one would not expect to find Matthew effects for all reading-related variables. Paris (2005) defined *constrained skills* as skills that are limited in scope, are learned quickly, and require the same material to be mastered by all learners, and argued that developmentally constrained skills "should not be conceptualized as enduring individual difference variables" (p. 184). Where outcome variables in other studies were constrained skills, such as word attack skills, one might not expect to find meaningful differences, especially for older or more skilled readers. Reading comprehension, on the other hand, is affected by different component skills through reading development. For very early readers, reading comprehension skill is largely a function of word reading or decoding skill. For more advanced readers, language comprehension skills make a more substantial contribution to reading comprehension. Because the components affecting reading comprehension scores differ in their contribution through development, longitudinal comparisons of reading comprehension skills may or may not show a Matthew effect. These challenges are compounded when combined measures of word reading and reading comprehension are used. Hence, it is possible that some previous studies have not found evidence to support the existence of a Matthew effect because the outcome measures were either developmentally constrained or were developmentally less constrained but were measured in age ranges before the effects would be expected to occur. This analysis would be supported by the meta-analysis of Pfost and colleagues (Pfost et al., 2014), which suggested that there was less evidence for a Matthew effect for developmentally constrained variables such as decoding accuracy. In the case of the current study, reading would be expected to affect vocabulary growth after children are exposed to a large number of novel words through reading, beginning at about the third or fourth grade. This is the developmental point investigated in this study.

Second, as discussed earlier, previous studies of a Matthew effect for vocabulary did not control for word-learning skills prior to formal reading instruction. Indeed, the results of the current study indicate that this variable has a significant effect on the rate of vocabulary growth in the years between the fourth and 10th grades. This may be another reason for the variable findings in previous studies.

Third, the current study used developmental ability scores based on IRT to allow for meaningful comparisons between performance at different age groups, which was not true of previous studies of a Matthew effect for vocabulary. The rationale for the use of IRT-based scores was that they appear to have the best properties, such as an equal-appearing interval scale, for characterizing the growth of mental abilities. Concerns have been raised as to whether these scores are likely to show declining variance with increases in age, whereas grade-equivalence scores seem to produce increasing variance (e.g., Hoover, 1984; Yen, 1986). These patterns, however, have not been consistently reproduced (Williams, Pommerich, & Thissen, 1998), and it remains unclear whether, or under what circumstances, IRT scores or other forms of developmental scores misrepresent the changes in ability over time. Nonetheless, the use of IRT (Rasch) scores has been critiqued in investigations of a Matthew effect fan spread (Bast & Reitsma, 1998; Stanovich, 2000) on the grounds that forcing within-age scores into a normal distribution could cause a decrease in developmental score variance with age (Hoover, 1984). For example, Stanovich (2000) proposed that the use of developmental ability scores could account for the compensatory effect found for reading scores in the study by Shaywitz et al. (1995). In the current study, the use of developmental ability scores did not result in a decline in variance, and thus did not prohibit our ability to detect an effect of word-reading ability on vocabulary growth. It may be that the results of Shaywitz et al. (1995) with respect to reading are due to the use of a reading composite score that includes word identification, pseudoword identification, and reading comprehension. As Bast and Reitsma (1998) also argued, a composite score with these three skills would not have comparable meaning over time, which significantly obscures the interpretation of the results.

Also, the current study used an epidemiologically based sample. This study was conducted with a group of children who came from a population sample and are therefore more diverse than are often found in research studies, especially where participants need to come into a laboratory setting. Thus, the findings of the present study are more likely to be representative of the population at large.

Relationship Between Reading and Vocabulary Growth Across Reading Skill Levels

The second question of this study was whether the relationship between reading skill and vocabulary growth was the same for both strong and weak readers. Indeed, further examination of the data revealed that the effect of early word-reading ability on vocabulary was not uniform across different levels of initial word-reading ability. Instead, it would appear that the strong readers made greater vocabulary gains relative to the average and weak readers. In the language of the Matthew effect, the rich were getting richer due to their better reading, but the poor were not getting poorer due to their weak reading. Morgan et al. (2008) also reported a Matthew effect that did not apply to both strong and weak readers although they reported asymmetry in the opposite direction, with students most at risk of reading disorders being more likely to fall behind in reading, whereas those least at risk not gaining with respect to typical readers. A different prediction of the Matthew effect model was being tested in this study, and this is likely to account for the difference in results.

In the current study, several factors might account for this one-sided Matthew effect, with a nonuniform effect of word-reading skill on vocabulary growth across skill levels. The current study does not differentiate between these possibilities and, naturally, they are not mutually exclusive. The first possibility is that the gap in reading volume between strong and average readers is greater than the gap in reading volume between average and weak readers. The possibility that there are larger differences in reading volume between strong and average readers, compared to the differences between average and weak readers, is somewhat speculative. However, Cunningham (2005) discussed data indicating that, for independent reading in fifth-grade students, the absolute differences between avid and average readers (90th and 50th percentiles for reading volume) are greater than are the absolute differences between average and weak readers (50th and 10th percentile for reading volume). This would be consistent with the hypothesis that differences in reading experiences are not in a linear relationship with skill level. In addition to the amount of reading in which individual children engage, it may also be that the reading material selected by strong readers contains a greater degree of novel vocabulary than does the material assigned to, or selected by, average or weak readers.

The second possibility is that students differ in the amount that they benefit from reading new words and that those individual differences are greatest between average and strong readers. There is evidence that children differ in their ability to derive word meanings from written contexts (Cain, Oakhill, & Elbro, 2003; Cain, Oakhill, & Lemmon, 2004; McKeown, 1985). Individual differences in word learning through text are related to differences in working memory and to the ability to learn new vocabulary in a direct instruction task (Cain et al., 2004). The existence of these individual differences motivates interventions to improve children's skill in deriving word meanings from context (Cain, 2007; Goerss, Beck, & McKeown, 1999; Nash & Snowling, 2006). Genetic evidence also provides support for the idea that environmental factors may have a nonuniform effect on vocabulary growth across skill levels. DeThorne, Petrill, Hayiou-Thomas, and Plomin (2005) reported that children with very low vocabulary scores had a higher heritability and a lower influence of shared environment, relative to children with less severe vocabulary deficits. Again, the possibility that these individual differences are greater between strong and average readers, compared to the differences between weak and average readers, is speculative.

It is also possible that weak readers were provided with educational interventions for reading skill or vocabulary knowledge, which reduced the cumulative disadvantage effect for them. The current study does not include information about intervention history, so this is speculative. However, this data suggests that the combination of behaviors chosen by students, differences in the ability to learn from exposure to new words in text, and educational policies are not further disadvantaging weak readers, at least in terms of their vocabulary growth. For those who are concerned about the poor getting poorer, this is an encouraging finding.

Limitations of the Current Study

As with any nonexperimental design, these conclusions are based on associations rather than on stronger experimental evidence involving random assignment to independent variable treatment conditions. The limitation of an observational design, such as this study, is that other confounding variables may play a role in observed effects. One such confounder could be the initial vocabulary level. Children with better word-reading skills in the fourth grade are also likely to have better listening vocabulary, as was true of the participants in this study. As a result, our analysis incorporated a measure of kindergarten-listening vocabulary ability as a covariate. This analysis is possible, in part, because the current study analyzes data from a large longitudinal sample of 485 participants, which provides adequate statistical power. We can assume that this measure of vocabulary in kindergarten was largely unaffected by the child's reading experience, but would reflect the child's general vocabulary–learning ability along with aspects of the child's environment that could be associated with individual differences in word learning. Thus, we can argue that the effects of fourth-grade reading ability on subsequent listening vocabulary are likely to be independent of the child's general vocabulary–learning ability.

Another limitation of the current study is the use of a linear model, which means that the vocabulary data between the fourth and 10th grades was fit using only a linear slope and intercept. However, the overall vocabulary scores between kindergarten and 10th grade suggest a curvilinear trend, and it is possible that the data might be better represented by a nonlinear model, but this would require data at more time points than is available with this data set. This means that the current analysis cannot address questions that are specific to acceleration or deceleration of vocabulary growth rate, but instead captures the primary feature of the growth trajectory, namely overall change through time.

The purpose of the original longitudinal study was to answer questions regarding outcomes of children with language impairment. The oversampling of children with language and/or cognitive impairments could potentially have biased the results of the current study. However, this was accounted for with the use of weighted scores. The findings, therefore, apply to both children and adolescents who were language impaired and typically developing.

Conclusions

The principal finding of this study is that fourth-grade reading-word skill was related to the rate of change in vocabulary growth between the fourth and 10th grades, controlling for preliterate vocabulary skill. We interpret measures of word reading in the fourth grade as being an indicator variable for a variety of reading-related activities occurring during and after the fourth grade, which would affect exposure to new words. The analysis controlled for vocabulary levels prior to formal reading instruction and used developmental scores based on IRT, addressing two potential limitations in studies of Matthew effects. Data in the current study was collected from a population-based sample, meaning that these findings apply to both readers who are typically developing and language impaired. Hence, the current study provides strong support for the existence of a Matthew effect between word-reading skill and vocabulary. It is significant that the magnitude of the effect on absolute vocabulary levels was found to be large. The effect seems to be driven by strong readers, rather than weak readers, an encouraging finding for those concerned about outcomes for weak readers. More broadly, these findings point to the importance of reading to the process of vocabulary acquisition in older children and adolescents.

References

Aarnoutse, C., & van Leeuwe, J. (2000). Development of poor and better readers during the elementary school. *Educational Research and Evaluation, 6*(3), 251–278.

Allington, R. L. (1983). The reading instruction provided readers of differing reading abilities. *The Elementary School Journal, 83*(5), 548–559.

Bast, J., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology, 34*(6), 1373–1399.

Biemiller, A. (2005). Size and sequence in vocabulary development: Implications for choosing words for primary grade vocabulary instruction. In E. Hiebert & M. Kamil (Eds.), *Teaching and learning new vocabulary: Bringing research to practice* (pp. 223–242). Mahwah, NJ: Erlbaum.

Biemiller, A., & Slomin, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology, 93*(3), 498–520.

Cain, K. (2007). Deriving word meanings from context: Does explanation facilitate contextual analysis? *Journal of Research in Reading, 30*(4), 347–359.

Cain, K., & Oakhill, J. (2011). Matthew effects in young readers: Reading comprehension and reading experience aid vocabulary development. *Journal of Learning Disabilities, 44*(5), 431–443.

Cain, K., Oakhill, J., & Elbro, C. (2003). The ability to learn new word meanings from context by school-age children with and without language comprehension difficulties. *Journal of Child Language, 30,* 681–694.

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology, 96*(4), 671–681.

Catts, H. W., Adlof, S. M., & Fey, M. C. (2003). Subgrouping poor readers on the basis of individual differences in reading-related abilities. *Journal of Learning Disabilities, 36*(2), 151–164.

Catts, H. W., Adlof, S. M., Hogan, T., & Ellis Weismer, S. (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech, Language, and Hearing Research, 48,* 1378–1396.

Catts, H. W., Fey, M. C., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading, 3*(4), 331–361.

Chall, J. S. (1987). Two vocabularies for reading: Recognition and meaning. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 7–17). New York, NY: Psychology Press.

Chou, C., Bentler, P. M., & Pentz, M. A. (1998). Comparisons of two statistical approaches to study growth curves: The multilevel model and the latent curve analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 5*(3), 247–266.

Cooter, R. B. (1989). Review of Woodcock Reading Mastery Tests–Revised. In J. Conoley & J. Kramer (Eds.), *The tenth mental measurements year book* (pp. 910–913). Lincoln, NE: Buros Institute of Mental Measurements.

Cunningham, A. (2005). Vocabulary growth through independent reading and reading aloud to children. In M. Kamil & E. Hiebert (Eds.), *Teaching and learning new vocabulary: Bringing research to practice* (pp. 45–68). Mahwah, NJ: Erlbaum.

DeThorne, L. S., Petrill, S. A., Hayiou-Thomas, M. E., & Plomin, R. (2005). Low expressive vocabulary: Higher heritability as a function of more severe cases. *Journal of Speech, Language, and Hearing Research, 48,* 792–804.

Dollinger, S. J., Matyja, A. M., & Huber, J. L. (2008). Which factors best account for academic success: Those which college students can control or those they cannot? *Journal of Research in Personality, 42,* 872–885.

Dunn, L. M., & Dunn, L. (1981). *Peabody Picture Vocabulary Test–Revised*. Circle Pines, MN: AGS.

Dunn, L. M., Dunn, L. M., Whetton, C., & Pintilie, D. (1982). *British Picture Vocabulary Scale*. Windsor, England: NFER-Nelson.

Gertner, B. L., Rice, M. L., & Hadley, P. A. (1994). Influence of communicative competence on peer preferences in a preschool classroom. *Journal of Speech and Hearing Research, 37,* 913–923.

Goerss, B. L., Beck, I. L., & McKeown, M. G. (1999). Increasing remedial student's ability to derive word meaning from context. *Journal of Reading Psychology, 20,* 151–175.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Brookes.

Hoover, H. D. (1984). The most appropriate scores for measuring educational development in the elementary schools: GE's. *Educational Measurement: Issues and Practices, 3*(4), 8–14.

Jacobson, T. C. (1999). How persistent is reading disability? Individual growth curves in reading. *Dyslexia, 8,* 78–93.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*(4), 437–447.

Kempe, C., Eriksson-Gustavsson, A. L., & Samuelsson, S. (2011). Are there any Matthew effects in literacy and cognitive development? *Scandinavian Journal of Educational Research, 55*(2), 181–196.

Martin-Chang, S. L., & Gould, O. N. (2008). Revisiting print exposure: Exploring differential links to vocabulary, comprehension, and reading rate. *Journal of Research in Reading, 31*(3), 273–284.

McKeown, M. G. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading Research Quarterly, 20*(4), 482–496.

Merton, R. K. (1968, January 5). The Matthew effect in science. *Science, 159*(3810), 56–63.

Mislevy, R., & Bock, R. D. (1998). Bilog 3 [Computer software]. Skokie, IL: Scientific Software.

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin, 137*(2), 267–296.

Morgan, P. L., Farkas, G., & Hibel, J. (2008). Matthew effects for whom? *Learning Disability Quarterly, 31,* 187–198.

Nagy, W., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly, 19*(3), 304–330.

Nagy, W., Herman, P., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20*(2), 233–253.

Nash, H., & Snowling, M. J. (2006). Teaching new words to children with poor existing vocabulary knowledge: A controlled evaluation of the definition and context methods. *International Journal of Language and Communication Disorders, 41*(3), 335–354.

Nelson, J. R., Michal, B., & Perfetti, C. (2005). Differences between written and spoken input in learning new words. *Written Language and Literacy, 8*(2), 101–120.

Newcomer, P., & Hammill, D. (1988). *Test of Language Development–Primary: Second Edition*. Austin, TX: Pro-Ed.

O'Malley, K. J., Francis, D. J., Foorman, B. R., Fletcher, J. M., & Swank, P. R. (2002). Growth in precursor and reading-related skills: Do low achieving and IQ discrepant readers develop differently? *Learning Disabilities Research and Practice, 17*(1), 19–34.

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly, 40*(2), 184–202.

Parrila, R., Auonola, K., Leskinen, E., Nurmi, J., & Kirby, J. (2005). Development of individual differences in reading: Results from longitudinal studies in English and Finnish. *Journal of Educational Psychology, 97,* 299–319.

Pearson, D. P., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly, 42*(2), 282–296.

Pfost, M., Dorfler, T., & Artelt, C. (2013). Student's extracurricular reading behavior and the development of vocabulary and reading comprehension. *Learning and Individual Differences, 26,* 89–102.

Pfost, M., Hattie, J., Dorfler, T., & Artelt, C. (2014). Individual differences in reading development: A review of 25 years of empirical research on Matthew effects in reading. *Review of Educational Research, 84*(2), 203–244.

Phillips, L. M., Norris, S. P., Osmond, W. C., & Maynard, A. M. (2002). Relative reading achievement: A longitudinal study of 187 children from first through sixth grades. *Journal of Educational Psychology, 94*(1), 3–13.

Protopapas, A., Sideridis, G. D., Mouzaki, A., & Simos, P. G. (2011). Matthew effects in reading comprehension: Myth or reality? *Journal of Learning Disabilities, 44*(5), 402–420.

Rigney, D. (2010). *The Matthew effect: How advantage begets further advantage*. New York, NY: Columbia University Press.

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence, 35,* 83–92.

SAS Institute. (2011). SAS Proc Mixed [Computer software]. Cary, NC: Author.

Scarborough, H. S., & Parker, J. D. (2003). Matthew effects in children with learning disabilities: Development of reading, IQ, and psychosocial problems from grade 2 to grade 8. *Annals of Dyslexia, 53,* 47–71.

Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., & Mermelstein, R. J. (2012). A practical guide to calculating Cohen's f2, a measure of local effect size, from PROC MIXED. *Frontiers in Psychology, 3*(111), 1–6.

Semel, E. M., Wiig, E. H., & Secord, W. (1996). *Clinical Evaluation of Language Fundamentals—Third Edition*. San Antonio, TX: The Psychological Corporation.

Shaywitz, B. A., Holford, T. R., Holahan, J. M., Fletcher, J. M., Stuebing, K. K., Francis, D. J., & Shaywitz, S. E. (1995). A Matthew effect for IQ but not for reading: Results from a longitudinal study. *Reading Research Quarterly, 30*(4), 894–906.

Shefelbine, J. (1990). Student factors related to variability in learning word meanings from context. *Journal of Reading Behavior, 22*(1), 71–97.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360–407.

Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. New York, NY: Guilford.

Sternberg, R. J. (1987). Most vocabulary is learned from context. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 89–106). Hillsdale, NJ: Erlbaum.

Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research, 41,* 407–419.

Tomblin, J. B. (2014). Background of the study. In J. B. Tomblin & M. A. Nippold (Eds.), *Understanding individual differences in language development across the school years* (pp. 1–11). New York, NY: Psychology Press.

Tomblin, J. B., & Nippold, M. A. (Eds.). (2014). *Understanding individual differences in language development across the school years*. New York, NY: Psychology Press.

Tomblin, J. B., Nippold, M. A., Fey, M. E., & Zhang, X. (2014). The character and course of individual differences in spoken language. In J. B. Tomblin & M. A. Nippold (Eds.), *Understanding individual differences in language development across the school years* (pp. 47–78). New York, NY: Psychology Press.

Tomblin, J. B., Records, N. L., Buckwater, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40,* 1245–1260.

Tomblin, J. B., & Zhang, X. (2006). The dimensionality of language ability in school-aged children. *Journal of Speech, Language, and Hearing Research, 49,* 1193–1208.

Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities, 45*(5), 453–466.

Vale, C. D. (1986). Linking item parameters onto a common scale. *Applied Psychological Measurement, 10*(4), 333–344.

Wallace, G., & Hammill, D. (1994). *Comprehensive Receptive and Expressive Vocabulary Test*. Austin, TX: Pro-Ed.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children–Third Edition*. San Antonio, TX: The Psychological Corporation.

Williams, V., Pommerich, M., & Thissen, D. (1998). A comparison of developmental scales based on Thurstone methods and item response theory. *Journal of Educational Measurement, 35*(2), 93–107.

Woodcock, R. (1987). *Woodcock Reading Mastery Test–Revised*. Chicago, IL: Riverside.

Yen, W. M. (1986). The choice of scale for educational measurement: An IRT perspective. *Journal of Educational Measurement, 23*(4), 299–325.